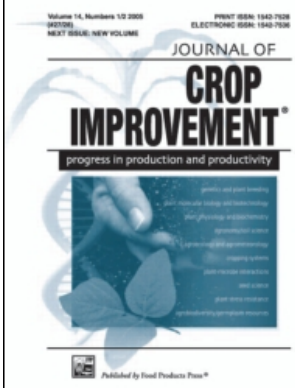


Journal of Crop Improvement
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t792303981

Usefulness of SSR Derived from Tetraploid Gossypium spp. for Analyses of Diploid Gossypium spp.

M. Altaf-Khan [a]; Samina N. Qureshi [a]; Sukumar Saha [b]; Johnie N. Jenkins [b]; Curt L. Brabaker [c]; Oumesh U. K. Reddy [d]
[a] Department of Plant and Soil Science, Mississippi State University, MS, USA [b] USDA, ARS, Crop Science Research Laboratory, MS, USA [c] CSIRO Plant Industry, Canberra, Australia [d] Molecular Biology Department, Alcorn State University, Alcorn, MS, USA

Online Publication Date: 20 July 2006

To cite this Article Altaf-Khan, M., Qureshi, Samina N., Saha, Sukumar, Jenkins, Johnie N., Brabaker, Curt L. and Reddy, Oumesh U. K.(2006)'Usefulness of SSR Derived from Tetraploid Gossypium spp. for Analyses of Diploid Gossypium spp.',Journal of Crop Improvement,16:1,1 — 20
To link to this Article: DOI: 10.1300/J411v16n01_01
URL: http://dx.doi.org/10.1300/J411v16n01_01

Usefulness of SSR Derived from Tetraploid *Gossypium* spp. for Analyses of Diploid *Gossypium* spp.

M. Altaf-Khan
Samina N. Qureshi
Sukumar Saha
Johnie N. Jenkins
Curt L. Brubaker
Oumesh U. K. Reddy

ABSTRACT. This investigation is the first attempt to utilize tetraploid species-derived primers to generate markers for mapping in diploid cotton species. About 113 individual F_2 plants, developed from the cross between *Gossypium nelsoni* and *G. australe,* were used to construct a linkage map. One allotetraploid (AD_1) *G. hirsutum* (TM-1) plant and its putative diploid ancestors, *G. arboreum* (A_2) and *G. raimondii* (D_5), were also included for comparative analysis. We developed a high-throughput multi-mix PCR assay for simple sequence repeat (SSR) marker separation using fluorescent-labeled capillary electrophoresis. Of a total of 205 tetraploid-derived SSR primers, 83% amplified in all four diploid species, indicating that flanking primer sequences are conserved in the diploid and tetraploid AD_1 genome of cotton. DNA sequence information of 18 randomly selected clones revealed that all of the clones contained

M. Altaf-Khan and Samina N. Qureshi are affiliated with Department of Plant and Soil Science, Mississippi State University, P.O. Box 9555, Mississippi State, MS 39762.

Sukumar Saha and Johnie N. Jenkins are affiliated with USDA, ARS, Crop Science Research Laboratory, P.O. Box 5367, Mississippi State, MS 39762.

Curt L. Brubaker is affiliated with CSIRO Plant Industry, G.P.O. Box 1600, Canberra, ACT 2601, Australia.

Oumesh U. K. Reddy is affiliated with Molecular Biology Department, Alcorn State University, Alcorn, MS 39096.

Journal of Crop Improvement, Vol. 16(1/2) (#31/32) 2006
Available online at http://www.haworthpress.com/web/JCRIP

doi:10.1300/J411v16n01_01

SSR repeat motifs in the diploid species, suggesting that SSR markers including the flanking primer sequences have been conserved at the inter-specific level in genomic DNA across years of evolution. About 85% of the amplified primers showed polymorphism among the diploid genome species. Forty percent of SSRs in the diploid *G. nelsoni* and 35% in *G. australe* amplified the same size DNA fragment as in the tetraploid, (TM-1). This is further evidence that many of the tetraploid SSRs are conserved across diploid and tetraploid species. *G. nelsoni* compared with *G. australe* had more conserved genomic regions in common with *G. hirsutum*. From 73 markers scored in the F_2 population, 50 exhibited codominance and 23 dominance. There were 47 markers linked across 10 linkage groups covering a total map distance of 754 cM with an average of 16 cM per marker. One linkage group (lg-3) with six loci has been assigned to chromosome 20. Comparative analysis of DNA sequences between *G. hirsutum* and its two putative parental diploid species indicated that changes within a few selected clones have occurred during evolution as follows: (1) by increasing or decreasing the number of the repeat motif, (2) by single nucleotide mutations occurring within or outside of the repeat motif, (3) by inter conversion of a simple to a compound SSR motif, and (4) by inversion of the repeat regions. Tetraploid-derived microsatellites should be useful in comparative genetic mapping of both tetraploid and diploid *Gossypium* species for evolutionary studies and introgression of agriculturally important traits from exotic diploid and tetraploid germplasm sources.

KEYWORDS. *Gossypium*, diploid species, simple sequence repeats (SSRs), linkage map, genome, evolution

INTRODUCTION

Cultivated species of *Gossypium* provide the world's leading natural fiber, cotton, and are also a major oil seed crop. The genus *Gossypium* includes approximately 45 diploid and five allotetraploid species. There are three major lineages of diploid species, corresponding to three continents: Australia (C, G, K genomes), the Americas (D genome), and Africa/Arabia (A, B, E, and F genomes) (Stewart et al., 1995; Percival et al., 1999). Upland cotton has a narrow genetic base and there is growing concern among cotton breeders that a lack of genetic diversity might limit future breeding advances.

The mapping of the cotton genome like other plant and animal species has been accelerated by the use of microsatellite markers. Genetic markers were developed in one of the A genome diploid cotton during the early 1900s, but few markers were developed for the two cultivated disomic tetraploids, *G. hirsutum* and *G. barbadense* (Endrizzi et al., 1985). There is an urgent need to explore the genetic diversity in natural populations and wild species of cotton. Much of the classification of existing germplasm collections of landraces and natural populations of wild *Gossypium* is based on morphological traits, which is an imprecise indicator of the genetic diversity. Hence their use in contemporary programs has sometimes met with failure, thereby stressing the importance of biochemical and molecular markers to evaluate populations derived from wild species to broaden genetic base of cotton.

Highly informative PCR-based genetic markers have been produced, which facilitate research in genetics, breeding, genome mapping, and sustainable management of genetic resources. These include amplified fragment length polymorphisms (AFLP) (Zabeau and Vos, 1993; Vos et al., 1995), and PCR-RFLP, also known as cleaved amplified polymorphic sequences (Konieczny and Ausubel, 1993), microsatellites or SSRs (Akkaya et al., 1992) or simple sequence length polymorphisms (SSLP) (Bell and Ecker, 1994). Microsatellites or SSRs are regarded as highly polymorphic genetic markers.

The SSR, a relatively new class of DNA markers, are highly informative genetic markers. Microsatellites are hypervariable sequence arrays of short motifs of 1-4 bp in length dispersed throughout plant genomes (Tantz, 1989; Weber and May, 1989; Wang et al., 1994). These single locus markers are characterized by their hypervariability, abundance, reproducibility, and Mendalian inheritance and are usually co-dominant in nature. The variation among microsatellites is thought to be due to the slippage of DNA polymerase during replication or unequal crossing over, resulting in differences in the copy number of the actual nucleotide sequences (Yu et al., 1999). Polymorphism among individuals arises from changes in the number of the repeats. In other words, these markers meet most of the requirements for ideal markers for assessing gene flow. Tracking of microsatellite markers requires specifically designed primers for the conserved flanking region of repeats and PCR amplification of this region.

Microsatellites are widely used in the construction of linkage maps, analysis of genetic structure of populations, and parent verification (Rafalski et al., 1996). They are also suitable for studies on the evolution of species (Goldstein and Pollock, 1997). A few examples of genetic

diversity research for *Gramineae* species include rice, *Oriza* spp. (Cho et al., 2000), finger millet (*Eleusine coracana*) (Salimath et al., 1995), maize, *Zea mays,* (Chin et al., 1996; Kantety et al., 1995), barley, *Hordeum* spp. (Becker and Heun, 1995; Russell et al., 1997), cotton (Liu et al., 2000), and wheat, *Triticum* spp. (Bryan et al., 1997; Eujayl et al., 2000; Plaschke et al., 1995).

Several SSR markers have been identified in tetraploid cotton species and successfully used to develop anchor SSR for cotton chromosomes to provide a basis for genetic mapping (Liu et al., 2000; Reddy et al., 2001). There are also reports that SSR primers specific to *G. hirsutum* L. have been used in mapping of *G. barbadense* L.(Liu et al., 2000a) and *G. nelsoni* L., and *G. australae* L. (Qureshi et al., 2001).

Cross-species transportability of SSR markers seems to be a good tool for genomic studies and can minimize the laborious cloning and screening steps of SSR development (Liu et al., 2000; Dayanandan et al., 1998, 1999; Echt et al., 1999). Similarly, microsatellite markers isolated from olive, *Olea europaea*, were used for fingerprinting of ancient cultivars (Cipriani et al., 2002).

The majority of linkage maps reported in cotton are based on RFLP markers (Reinisch et al., 1994; Shappely et al., 1998; Yu et al., 1998; Brubaker et al., 1999; Ulloa et al., 2000, 2002; Brubaker et al., 2000). Because of polyploidy and the complexity of tetraploid cotton genetics, it is laborious and time consuming to develop a saturated RFLP map of this large genome. An efficient PCR-based SSR system should be useful. However, there is no report on the use of SSR markers from tetraploid cotton in the mapping of diploid cotton species. The objectives of the present study were (1) to determine if tetraploid-derived SSR primers can be used to amplify DNA fragments in four diploid genomes, (2) to use tetraploid-derived SSR to construct a linkage map of an F_2 population from parental diploid G genome species (*G. nelsoni* and *G. australe*), (3) to characterize the amplified products from diploids for similarity of repeat motifs of SSR in the tetraploid.

MATERIAL AND METHODS

Plant Material

An F_2 population of 113 individual plants was developed from a cross between two Australian diploid species of cotton (*G. nelsoni* and *G. australe*) and used for genetic analysis and map development. TM-1, the

genetic standard for Upland allotetraploid cotton genetics (*G. hirsutum*-AD_1), was included for comparative purposes. The diploid species *G. arboreum* (A_2) and *G. raimondii* (D_5), which are regarded as the putative ancestors of Upland cultivated cotton, were also included in the study.

DNA Extraction

DNA samples from the two diploid G species, the F_1, and 113 individual F_2 plants were provided by Curt L. Brubaker (CSIRO, Australia). Genomic DNA of A_2, D_5, and TM-1 plants was isolated at the micro-level from young leaves using a CTAB-based extraction method of Altaf et al., (1997) with slight modifications. Approximately 0.5 g of fresh young leaf tissue was homogenized in 0.7 ml of extraction buffer [100 mM tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), 1 M NaCl, 2% CTAB, 2% PVP-40, 1 mM 1-10, phenanthroline] and 0.2% β-mercaptoethanol in a 1.5 ml centrifuge tube with the aid of a microtube pellet pestle and an electric hand drill. After incubation for 1 hr at 60°C, the suspension was purified twice in chloroform:isoamyl alcohol (24:1) by centrifugation at 10,000 rpm on a desktop micro-centrifuge for 10 min at room temperature and precipitated with an equal volume of cold isopropanol. The recovered DNA was spooled out, or pelleted by centrifuging at 10,000 rpm for 5 min, washed twice with 80% EtOH + 15 mM ammonium acetate and once with 95% EtOH, air dried, and dissolved in 100 to 200 µl of 10 mM Tris buffer (pH 7.5). To this sample 2 µl of Rnase A (10 mg/ml) per 100 µl of dissolved DNA was added.

Microsatellite PCR Amplification

Cotton-specific SSR primers were purchased from Research Genetics (Huntsville, AL) and fluorescent labeled by Perkin-Elmer, Applied Biosystems. Multiplex PCR conditions were used, as per the method of Liu et al. (2000), with some modifications. PCR reactions were performed in 10 µl volume containing 50 ng of cotton DNA, 1X Perkin-Elmer PCR BufferII (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 2.5-3.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.15 mM of each single primer, 0.4 units of AmpliTaq Gold (Perkin-Elmer Applied Biosystems). The amplification profile consisted of an initial period of DNA denaturation and AmpliTaq Gold activation at 94°C for 7 min, followed by 9 cycles (step 1) of 94°C for 15 sec, 65°C for 30 sec and 72°C for 1 min, by decreasing annealing temperature 1°C in each cycle from 65°C to 56°C and 40 cycles (step 2) of

94°C for 15 sec, 55°C for 30 sec and 72°C for 1 min. The extension temperature of 72°C was held for 7 min. To verify amplification, PCR products were run on 2% agarose gel (Figure 1).

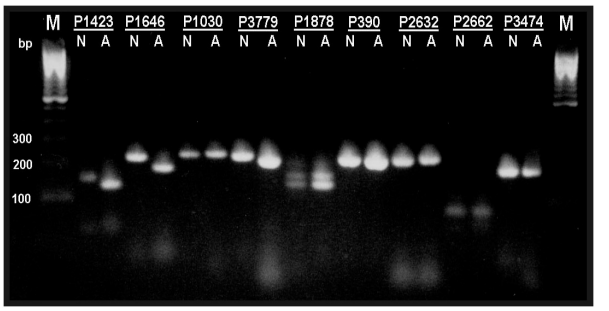
Gel Electrophoresis

Multiplex PCR products were separated and detected by using the ABI PRISM 310 Genetic Analyzer (Perkin-Elmer, Applied Biosystems). PCR products were diluted 1:20 for HEX and FAM, 2:20 for NED labeled primers with dd H_2O. An aliquot of 3 µl of diluted solution and 0.2 µl of ROX dye (an internal size standard) was added to 10 µl of formamide, denatured at 95°C for 5 min, kept at 4°C for 5 min and loaded into ABI PRISM 310 Genetic Analyzer (Perkin-Elmer, Applied Biosystems). The amplified products were quantified using Gene Scan software.

Sequencing of PCR Products

To confirm the presence of the microsetellite repeats in the amplified fragments, 18 amplified SSR PCR products were cloned and sequenced. SSR products were amplified using selective unlabeled primer pairs and run on 2% agarose gel for confirmation of right amplification products. Cloning was performed using PCR4-TOPO TA Cloning® Kit for sequencing (Invitrogen Corporation, Carlsbad, CA) following the standard vendor instructions. Sequencing of these clones was carried out via BigDye terminator cycle sequencing system (Perkin-Elmer, applied Biosystems, Foster City, CA) using M13-forward or M13-reverse primer.

FIGURE 1. Amplified products of a few selected SSR markers of diploid parent species (G genome) were isolated using agarose gel and sequenced for comparative analysis with tetraploid cotton. M = 100 bp DNA marker, P = SSR primer #, N = *G. nelsoni*, A = *G. australe.*



Electrophoretic separation of sequencing products was performed on an ABI PRISM 3700 DNA Analyzer 96-capillary automated sequencer (Perkin-Elmer, Applied biosystems).

Sequence Analysis

Base calling was performed using Chromas ver. 1.5 (Technelysium Pty. Ltd. Queensland, Australia). Contigs of double-stranded DNA sequences from each clone, identification of overlaps, redundancy among clones, and vector sequence trimming were performed via Sequencher ver. 4.0.5 (Gene Codes, Ann Arbor, MI). Amplified SSR sequences of *G. hirsutum* (TM-1) were obtained from the cotton database at (http://algodon.tamu.edu/htdocs-cotton/cottondb.html) and sequences were aligned using web-based Clustal W 1.8 program. The dendrogram of DNA sequences was constructed via MegAlign program of DNAStar software package.

Linkage Analysis and Map Construction

All segregating markers in the F_2 population were subjected to χ^2 test at 0.1% probability. Linkage analysis was performed using MapMaker 3.0 set at a LOD score ranging from 3.0 to 5.0 and a maximum distance of 50 cM between linked markers using Kosambi unit. Markers were ordered on linkage groups by using Compare and Order commands. Graphical diagrams of linkage groups were drawn by using web-based software GGT 32.

RESULTS AND DISCUSSION

The Upland cotton genome (AD) has a large amount of repetitive DNA similar to other eukaryotic plant genomes (Endrizzi et al., 1985; Baker et al., 1995; Zhao et al., 1998). Physically, the D genome chromosomes are shorter than the A genome chromosomes, but a greater amount of repetitive DNA is present in the A genome than in the D genome (Endrizzi et al., 1985). The regions flanking a microsatellite are highly conserved at both the intra- and inter-specific levels in tetraploid cotton and PCR primers for the flanking regions were used to identify chromosomal location of many of the available SSRs (Liu et al., 2000). Several researchers have used the SSR method for genetic analysis in cotton because of their highly polymorphic nature (Altaf et al., 1997; Feng et al.,

1997; Liu et al., 2000). They have been used in marker-assisted selection (Tanksley and McCouch, 1997) of wild and primitive germplasm resources of other crops.

The cotton genome is genetically very diverse, about 5000 cM in size. Several thousand microsetellites will be required to develop a high-resolution saturated genetic map of Upland cotton. More than 500 microsatellite clones, containing primarily (GA)n repeats, have been identified at Brookhaven National Laboratory. Several of these sequences are redundant. Primer pairs for the amplification of ~240 of these loci (designated BNL for Brookhaven National Laboratory) have been made available to the cotton research community through purchase from Research Genetics, Huntsville, AL. An additional 150 (GA)n repeat loci (designated CM for cotton microsatellite) have been isolated at Texas A & M University (Connell et al., 1998; Reddy et al., 2001). Reddy et al. (2001) identified an additional 300 SSR markers using SSR-enriched genomic libraries. Liu et al. (2000) reported the chromosomal assignment of several BNL and CM microsatellite markers in Upland cotton using cytogenetic stocks. SSR primers specific to *G. hirsutum* L. have been used in mapping *G. barbadense* L.(Liu et al., 2000). Despite these advances, a large number of additional markers are needed to meet the needs of functional genomic mapping in cotton.

Use of Tetraploid-Derived SSR Primers in Diploid Species

The utility of microsatellite markers for the study of genetic diversity and mapping within wild and cultivated germplasm of some crop plants is well established. Due to the narrow genetic base of cultivated cotton, there is a need to use genes from wild germplasm to meet future genetic challenges. Regions flanking microsatellite loci are often conserved among closely related species (Moore et al., 1991), allowing the use of primers developed for one species to amplify loci in closely related species. Although these SSR primers were derived from tetraploid *G. hirsutum* species, they have successfully amplified microsatellite flanking regions of two wild diploid G genome species and the two wild diploid putative ancestor species (*G. arboreum* and *G. raimondii*) of Upland cotton.

A total of 255 primers derived from tetraploid upland cotton species were used to amplify genomic DNAs of diploid *G. nelsoni* and *G. australe* (G genome), their F_1 and individual F_2 plants. DNA samples from tetraploid *G. hirsutum* genetic marker line TM-1 (AD_1 genome) and one of its diploid ancestor *G. raimondii* (D_5 genome) were also used for the

comparison of amplified products between tetraploids and diploid species. Seventeen percent of the total primers analyzed did not give any amplification among the five species analyzed in this study and 19% of the primers did not give any amplification product among any diploid genome species. Although the G genome species (*G. nelsoni* and *G. australe*) used in this study originated from Australia and are distantly related to cultivated tetraploid cotton, they showed a high level of amplification and only a small portion of primer pairs failed to amplify a product. These failed amplification products are most likely due to sequence alterations, such as point mutations, deletions or inversions, within the priming sites as reported by Devos et al. (1995).

About 85% of the tetraploid-derived primers that produced amplified products exhibited polymorphism among the four diploid species. The polymorphic primers between the two diploid G genome parents amplified 1-3 DNA markers per primer. This indicates a great genetic diversity exists at the molecular level between these two morphologically similar diploid species of the G genome. The trend of each SSR primer to amplify 1-2 markers in diploid species is similar to what Liu et al. (2000) reported for use of primers across two tetraploid species. We also observed that amplified SSR product always showed stutter bands, which is a typical pattern for SSR markers. This provides further evidence that the amplified band is an SSR product. Our results are the first report of SSR markers in diploid G cotton species, and they also provide evidence that many of the currently available SSR tetraploid-derived primers can be used to map diploid species of cotton.

Comparisons Among Tetraploid Upland Cotton and Its Diploid Ancestors

The cultivated species of cotton, *G. hirsutum* (AD_1 genome) is a tetraploid and has reportedly evolved from hybrids of two diploid species, *G. arboreum* (A_2) and *G. raimondii* (D_5) over a long period of time. In this study, we also included three important genomic groups (A, D, AD_1 genomes) in addition to the two G genome species (*G. nelsoni* and *G. australe*). Comparison of homologous genome-specific repetitive DNA sequences from closely related organisms not only provides information about the evolution of the sequences themselves, but also shows relationships among species.

The present examination of tetraploid-derived SSRs in diploid and tetraploid species revealed some interesting features. When we compared the PCR-amplified products produced from the same set of SSR

primer pairs across the genomes (Table 1), we observed some very important evolutionary aspects of ploidy formation. We found that 43% and 42% of the amplified SSR products from A_2 and D_5 genomes, respectively, were of the same size as the amplified product in the tetraploid genome AD_1. We found 40.5% and 33.3% of the amplified products from the two G genome diploids were of the same size as the SSR products of the tetraploid AD_1. If we compare the same products in both G genomes with any of the other three genomes, however, there was less than 10% size homology. We found 43.1% of the SSR fragments were of the same size between A_2 and D_5 genomes, indicating a close relationship between these two diploid species. When we compared the G genome, we found that 40.5% of *G. nelsoni* fragments were of the same size as in the AD_1 genome; whereas 33.3% of the *G. australe* fragments were of the same size as the AD_1 fragments. A slightly greater percentage of the amplified fragments from *G. australe* were size-monomorphic with A_2 and D_5 than when *G. nelsonii* was compared with A_2 and D_5 (Table 1). This study revealed that a considerable number of copies of repetitive DNAs are conserved among closely and distantly related species of the genus *Gossypium* and these cross-specific regions of the genome can be utilized for comparative mapping to expand the genetic base of cotton.

Inheritance of G Genome Markers

In the F_2 population, dominant and co-dominant markers were scored based on parental screening of *G. nelsoni* and *G. australe* (Table 2). We scored 73 polymorphic SSR markers in the F_2 population. Out of these 73 markers, 50 exhibited co-dominance and 23 dominance in F_2 population. Chi-square test indicated that 44% of the markers exhibited distorted segregation in the F_2 population. Up to 70% distorted segregation has been reported in diploid populations of alfalfa (Diwan et al., 2000)

TABLE 1. Percentage of SSR markers homology among tetraploid and diploid genomes of cotton.

Genome	AD_1	A_2	D_5	G
G. nelsoni	40.5	20.2	21.7	
G. australe	33.3	35.4	29.1	
AD_1		43.1	42.2	8.5
A_2			43.1	5.1
D_5				6.8

TABLE 2. Inheritance and segregation of SSR markers among F_2 population, parents and primer sets.

Genotypes and Primer Sets	Inheritance		Segregation	
	Co-Dominant	Dominant	Normal	Distorted
F_2	50	23	41	32
G. nelsoni		9	2	7
G. australe		14	5	9
BNL primers	29	19	31	17
JESPER primers	21	4	10	15

and many other crops, including cotton (Reinisch et al., 1994; Shappely et al., 1998; Brubaker et al., 1999, Altaf et al., 1999; Ulloa et al., 2002). Out of 23 dominant markers, 14 originated from *G. australe* and 9 from *G. nelsoni* with a distortion percentage of 39% and 30%, respectively, within dominant markers for these two species. Of the 73 polymorphic markers segregating in the F_2 population, 48 markers were amplified by the BNL primer pairs and 25 markers from JESPER primer pairs and these had distorted segregation in 35% and 60% of the markers, respectively (Table 2).

Mapping of the Diploid G Genome

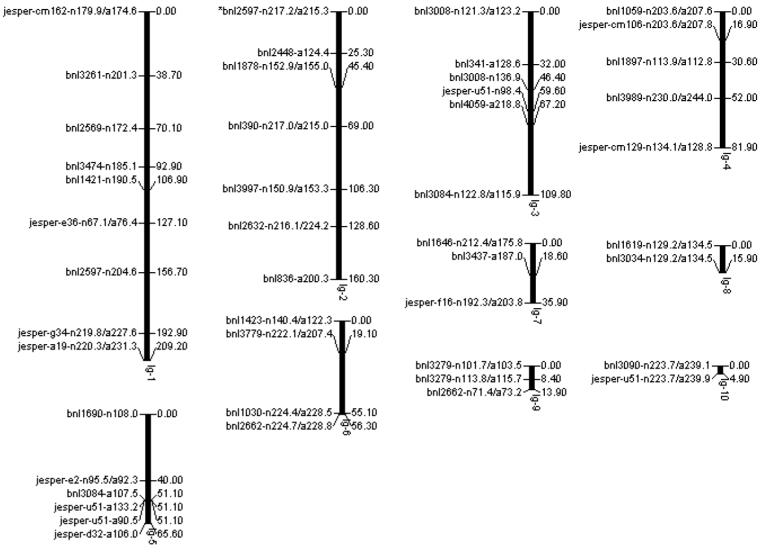
Many useful applications of molecular markers require a genetic map that is reasonably saturated with markers. The array of mapped markers now available in some crops is bringing such applications within the reach of researchers. Even the more demanding and sophisticated applications (e.g., QTL mapping, positional cloning, marker-assisted selection) are now materializing in crops such as tomato (*Lycopersicum esculentus*), maize, rice, lettuce (*Lactuca sativa*), and potato (*Solanum tuberosum*) as well as *Arabidopsis* spp. Despite these advances in many plant species, a single, high density, unified map for cotton does not exist. In cotton, existing maps have been developed mostly using RFLP marker systems (Reinisch et al., 1994; Shappely et al., 1998; Yu et al., 1998; Brubaker et al., 1999; Ulloa et al., 2000, 2002; Brubaker et al., 2000).

Only two maps comprising most of the AFLP markers in cotton (Altaf-Khan et al., 1999; Reddy et al., Unpublished) have been reported up to date. Until now, the genomic information in the cotton genome has not been explored via use of SSR markers due to the paucity of available SSR primers in cotton. Our diploid map is the first reported cotton map

among the G genome species that uses SSR markers. In addition, the map provides the opportunity to explore new genetic avenues to use the Australian G genome diploid species to expand the narrow genetic base of the Upland cotton.

In this study 73 polymorphic markers were scored from the F_2 population of *G. nelsoni* × *G. australe* and used to develop a linkage map (Figure 2). Out of 73 markers, 47 were linked across 10 linkage groups, covering a total map distance of 753.7 cM with an average of 16 cM distance between markers, while 26 markers were unlinked. The largest linkage group (lg-1) had a map distance of 209.2 cM with 9 markers; whereas, the smallest linkage group had a map distance of 4.9 cM with 2 markers. In linkage group one (lg-1), there were 4 heterozygous loci generated by the JESPER primer set and the remaining 5 loci were generated by BNL primer set and showed dominant inheritance with the BNL markers originating from female parent *G. nelsoni*. It appears that the genomic contribution in this linkage group was more from female

FIGURE 2. Linkage groups developed from F_2 population of *G. nelsoni* and *G. australe*. Marker name is written on the left side that includes SSR primer name (either BNL or JESPER), primer number, origin of marker "n" for *G. nelsonii* and "a" for *G. australe* followed by marker size in bp. Genetic distanc in cM is on right side and name of each linkage group is written at the bottom.



parent than male parent. Linkage group 2 (lg-2) had 5 heterozygous markers and 2 dominant markers from male parent (*G. australe*). Linkage groups lg-3, lg-4, lg-6, lg-8, lg-9, and lg-10 possessed mostly heterozygous loci and showed approximately an equal contribution of markers from both parents. Inheritance of lg-5 and lg-7 was predominantly from male parent as a majority of the dominant loci originated from *G. australe*.

Markers were randomly and evenly distributed over the 10 linkage groups of the G genome map except for lg-5 and lg-6. In lg-5, three dominant loci (BNL3084-a107.5, JESPER-u51-a133.2, JESPER-d32-a106.0) from *G. australe* were clustered at one point and in lg-6 two heterozygous loci (BNL1030-n224.4/a228.5 and BNL2662-n224.7/a228.8) had less than 1 cM distance between them. This indicates a lack of recombination at these points and these locations might be in the centromeric or telomeric regions. One linkage group (lg-3) has been assigned to chromosome 20 of the cotton genome based on information from Liu et al. (2000) about anchor SSR loci. In this linkage group, two loci (BNL3008-n136.9 and BNL 3008-n121.3/a123.2) were generated by the same primer pair (BNL3008) and in the same product size range as reported by Liu et al. (2000).

Diploid cotton genome has 26 chromosomes (n = 13, 2x = 26) with an approximate genome size of 2000-2500 cM; whereas Upland cotton genome size is estimated at 5000 cM (Reinisch et al., 1994). Our diploid G genome map covered ~35% of the cotton diploid genome. These mapped SSR markers should aid in the identification of chromosomal regions in the A and D genomes of tetraploid *Gossypium* species that are homeologous to the diploid species of the G genome and should thus accelerate gene identification efforts based on positional cloning and candidate gene approaches. Reinisch et al. (1994) discussed the advantages of using diploids for physical mapping and map-based cloning to overcome the challenges due to complex nature of the polyploid Upland cotton (2n = 26, 4x = 52).

Repeat Type and Sequence Homology of SSR-Amplified Products Between Diploid G Genome and the Tetraploid AD_1 Genome

In order to demonstrate the presence of bonafide teraploid SSR ortholoci in these wild diploid species, PCR products were cloned from amplifications using primer pairs that showed substantial cross-species transfer. A total of 18 PCR products (7 from *G. nelsoni* and 11 from *G. australe*) amplified by 11 primer pairs were cloned and sequenced to

confirm the presence of repeat motifs in these products (Table 3). These products included co-dominant and dominant markers from both species. The repeat motifs for tetraploid *G. hirsutum* were obtained from cotton database for the same primer pair and compared with the repeat motifs we identified in the diploid species. Both diploid species, *G. nelsonii* and *G. australe*, contained similar perfect $(CT)_n$, $(GA)_n$ and compound $(CA, TA)_n$ repeat types in both 5′-3′ and 3′-5′ directions except BNL 3997 that generated perfect repeats in *G. nelsoni* but compound repeats in *G. australe*. Of the two dominant markers from *G. australe,* one (BNL2569) exhibited perfect $(GA)_8$ and the other (BNL-3474) showed interrupted (CA_5, AA, CA_9) repeat types. There were considerable similarities and differences between repeat types in tetraploid *G. hirsutum* and diploid species. These differences might be due to the genomic distance between the tetraploid and diploid species as well as mismatches during DNA amplification process. Repetitive elements in the genome vary considerably among species and the Upland cotton genome appears to be among those containing the fewest families and the lowest relative abundance of tandem repeated DNA (Baker et al., 1995). Walbolt and Dure (1976) observed that the cotton genome was comprised of approximately 60.5% unique sequence DNA, 27% middle repetitive sequence DNA and the remaining approximately 12.5% highly repetitive sequence DNA. Baker et al. (1995) also found

TABLE 3. Selected SSR primer products after sequencing with amplified marker size and repeat motif among three species.

Primer	Marker Size (bp)		Repeat Motif		
	G. nelsoni	*G. australe*	*G. nelsoni*	*G. australe*[*]	*G. hirsutum*
BNL390[1]	–	215	–	GA_5	GA_{11}, T, AG_5
BNL1030	224	228	CA_5, TA_{11}	CA_7, TA_{10}	GT_{16}, CA_{13}
BNL1423	140	122	CT_{18}	CT_{13}	AG_{12}
BNL1646	212	175	CT_{18}	CT_{19}	AG_{20}
BNL1878[!]	152	155	GA_6	CT_6	AG_{14}
BNL2569	–	155	–	GA_8	GA_{13}
BNL2632	216	224	GA_{23}	GA_{23}	GA_{14}, TC_{14}
BNL3474[!]	–	187	–	CA_5, AA, CA_9	CA_{16}
BNL3779	222	207	CA_7, TA_4	CA_7, TA_5	TG_{13}
BNL3989[1]	–	244	–	CT_{25}	TC_{16}
BNL3997	150	153	CA_{17}	CA_8,CT_3	AC_{19}, TG_{19}

[1] Only one fragment was sequenced.
[*] Repeat motif were obtained from cotton database.
[!] Dominant marker only.

similar results, observing that the upland cotton genome consists of approximately 61% unique and low copy number DNA sequences, 32% highly repetitive and middle repetitive classes of DNA and 3% highly repetitive class of DNA sequences. They found a higher percentage of $(CT)_n$ and $(GT)_n$ dinucleotide repeat types compared with other types. Dinucleotide and trinucleotide repeat types are widespread in the cotton genome if one considers several species. Reddy et al. (2001) targeted SSR regions in cotton to develop a new set of SSR primers called "JESPER primers" and found 32% dinucleotide, 42% trinucleotide, and 1.3% hepta- and hexanucleotide repeat types. In their study, 25% of the SSR comprised compound repeat types with the most abundant being the AGA repeat, followed by GA, CA, and ACA repeat types. Our data from cloned sequences of diploid species *G. nelsoni* and *G. australe* consisted of the same repeat types as present in the tetraploid species of the cotton genome. The predominant SSR motif isolated in this study was of the types $(CT)_n$ and $(CA)_n$ (Table 3).

Figure 3 illustrates a DNA sequence alignment for locus BNL1030, using a single representative clone from each diploid species. This cloned sequences from PCR products of the diploid species showed a very high degree of identity for sequence and repeat type with a sequence of the same primer product from the tetraploid cotton obtained from the cotton database. Some clone-to-clone variation was observed among sequences derived from both diploid species, which may be attributable to primer mismatch and or allelic variation due to heterozygosity.

DNA sequence analysis of individual cloned amplification products confirmed the presence of SSR repeats in both of the wild diploid species. A high degree of sequence conservation was detected among the flanking regions of diploid and tetraploid species, although single nucleotide differences and various pairwise insertion or deletion polymorphisms distinguished each of the species.

Phylogeny of SSR Flanking Regions Among Diploid and Tetraploid Genomes

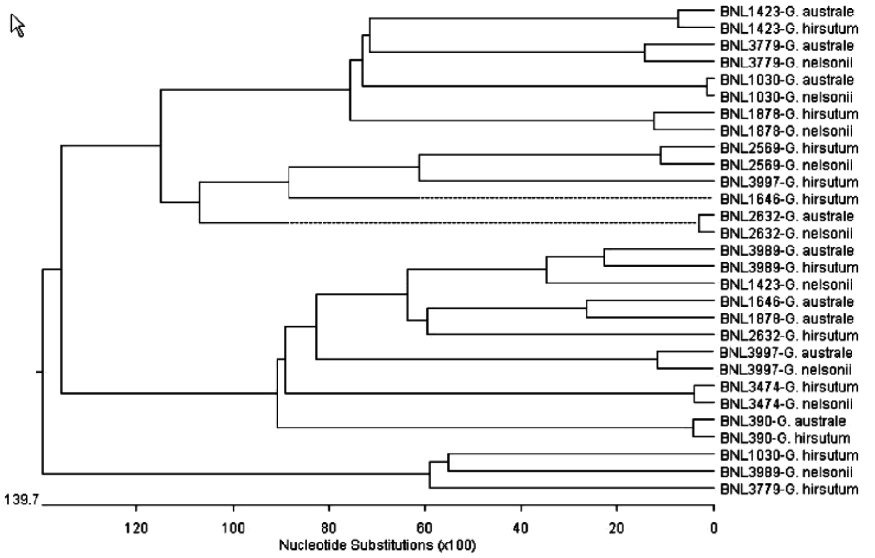
A dendrogram was generated from 18 cloned DNA sequences of *G. nelsoni* and *G. australe* and DNA sequences of *G. hirsutum* from the cotton database for the same SSR primer pairs (Figure 4). The combined data generated by flanking regions of 11 primer pairs distinguished three species at DNA sequence level. Four types of phylogenic cluster combinations were observed. First, four primer pairs (BNL3779, BNL1030, BNL2632, BNL3997) produced closely related homologous

FIGURE 3. Sequence alignment of three species from cloned amplified products using one teraploid derived SSR primer pair (BNL1030). DNA sequence for *G. hirsutum* was obtained from cotton database. DNA sequence identities across three species are indicated with an asterisk (*). Regions with repeat motifs are represented under the bold line.

```
BNL1030-G.hirsutum      ----------------------GAGAGAGAAACCACTTCTGCATCTGGATCTAACATCAC
BNL1030-G.nelsonii      ------------------------------------------------GGATCTAACAGCAC
BNL1030-G.australe      GTCCTGCAGGTTTAAACGAATTCGCCCTTAAACCACTTCTGCATCTGGATCTAACAGCAC
                                                                **********  ***

BNL1030-G.hirsutum      CCAATTATATTTTCCAATTAGACAAAAAGAAAGCCAATCATGAACACAAACACACACACA
BNL1030-G.nelsonii      CCAATTATATTTTCCAATTAGACAAAAAGAAAGCCAATCATGAACACA------------
BNL1030-G.australe      CCAATTATATTTTCCAATTAGACAAAAAGAAAGCCAATCATGAACACA------------
                        **************************************************

BNL1030-G.hirsutum      CACACACACACACACATATATCCATTTTGAATA----ATTGGGTAAAAAGCTTTCAT--A
BNL1030-G.nelsonii      ------CACACATATATATATCCATTTTGGATATATAATTGGGTAAAAAGCTTTCATTTA
BNL1030-G.australe      --CACACACACATATATATATCCATTTTGGATATATAATTGGGTAAAAAGCTTTCATTTA
                              ****** * *************** ***    **********************  *

BNL1030-G.hirsutum      TATATATATAATACGTACCAGGTGTTCCTGCGGGTCTTCTTTTCCTCTTATTGTTGTTAT
BNL1030-G.nelsonii      TATATATATAATACGTACCAGGTGTTCCCGCGGGTCTTCTTTTCCTCTTATTGTTGTTAT
BNL1030-G.australe      TATATATATAATACGTACCAGGTGTTCCCGCGGGTCTTCTTTTCCTCTTATTGTTGTTAT
                        ***************************** ********************************

BNL1030-G.hirsutum      TACCATTGTCTATGCATGTAAATGGCTCCAAAGAAGGAAGTGAGG--
BNL1030-G.nelsonii      TACCATTGTCTATGCATGTAAATGGCTCCAAACTGGGCGGTTTTATG
BNL1030-G.australe      T-----------------------------------------------
                        *
```

sequences with less than 20 bp substitutions between *G. nelsoni and G. australe*, indicating the conservation of this region between these two species. Second, three primer pairs (BNL1878, BNL 2569, BNL-3474) exhibited homology between *G. hirsutum* and *G. nelsoni* with a variation of less than 20 nucleotides. Third, another set of three primer pairs (BNL390, BNL1423, BNL3989) also targeted a conserved region between *G. hirsutum* and *G. australe* with a range of nucleotide substitution from ~5-25 bp. Fourth, there were four major groups of these DNA sequences among three species with a varying numbers of nucleotide substitutions revealing that SSR flanking regions contained a high number of informative substitutions and such variation proved to be particularly useful for distinguishing between very closely related species, such as *G. nelsoni* and *G. australe*. The resolving power of this analysis is clearly limited by the number of species and number of cloned sequences used; nevertheless, the main aim of this investigation was not to define phylogenetic relationships among these species but rather to show the potential application of SSR flanking regions across the species.

FIGURE 4. A phylogenetic dendrogram developed by DNA sequences SSR flanking regions among *G. hirsutum*, *G. nelsoni*, and *G. australe*. DNA sequences for *G. hirsutum* were obtained from cotton database.



In conclusion, benefits of investigating cross-species amplification of SSR loci in the *Gossypium* are multiple. These studies indicated that tetraploid-derived SSR primers could be successfully used in the distantly related diploid species of cotton. Large portions of SSR flanking regions appear to be conserved among diploid and tetraploid genomes of cotton. Linkage groups developed in this study are useful in comparative genomic and evolution analysis between tetraploid and diploid species because many of these markers have been mapped in the tetraploid species of cotton.

REFERENCES

Akkaya, M. S., A. A. Bhagwat, and P. B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean genetics. Genetics 132:1131-1139.

Altaf, M. K., J. McD. Stewart, M. K. Wajahatullah, J. Zhang, and R. G. Cantrell. 1997. Molecular and morphological genetics of a trispecific F_2 population of cotton. In P. Dugger and D. Richter (Eds.) Proc. Beltwide Cotton Production Research Conferences. Jan. 6-10, New Orleans, LA. National Cotton Council of America, Memphis, TN. Pp 448-452.

Altaf Khan, M., G. O. Myers, J. McD. Stewart, J. Zhang, and R. G. Cantrell. 1999. Addition of new markers to the trispecific cotton map. In P. Dugger and D. Richter (Eds.) Proc. Beltwide Cotton Production Research Conferences. Jan. 3-7, Orlando, FL. National Cotton Council of America, Memphis, TN. Pp 439.

Baker, J. and Heun, M. 1995. Barley microsetellites: allelle variation and mapping. Plant Mol. Biol. 27:835-845.

Baker, R. J., J. L. Longmire, and R. A. Van Den Bussche. 1995. Organization of repetitive elements in the upland cotton genome (*Gossypium hirsutum*). J. Hered. 86:178-185.

Bell, C. J., and J. R. Ecker. 1994. Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. Genomics 19:137-144.

Brubaker, C. L., A. H. Paterson, and J. F. Wendel 1999. Comparative genetic mapping of allotetraploid cotton and its progenitors. Genome 42:184-203.

Brubaker, C., E. R. Cantrell, M. Giband, B. Lyon, and T. Wilkins. 2000. Letter to Journal of Cotton Science community. J. Cotton Sci. 4:149-151.

Becker, J., Vos, P., Kuiper, M., Salamini, F., and M. Heun, 1995. Combined mapping of AFLP and RFLP markers in barley. Mol. Gen. Genet. 249:65-73.

Chin. B. C. L., M. L. Senior, H. Shu., and J. S. C. Smith. 1996. Maize simple repetative DNA sequences: aboundance and allele variation. Genome 39:866-873.

Cho, Y. G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S. R. McCouch, W. D. Park, N. Ayres, and S. Cartinhour, 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100: 713-722.

Connell, J. P., S. Pammi, M. J. Iqbal, T. Huizinga, and A. S. Reddy. 1998. A high through-put procedure for capturing microsatellites from complex plant genomes. Plant Mol. Biol. 16:341-349.

Devos, K. M., G. J. Bryan, A. J. Collins, P. Stephenson, M. D. Gale. 1995. Application of two microsatellite sequences in wheat storage proteins as molecular markers. Theor. Appl. Genet. 90:247-252.

Diwan, N., J. H. Bouton, G. Kochert, and P. B. Cregan. 2000. Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. Theor. Appl. Genet. 101:165-172.

Endrizzi, J. E., Turcotte, E. L., and Kohel, R. J. 1985. Genetics, cytology, and evolution of *Gossypium*. Advances in Genet. 23:271-375.

Eujayl, I., Sorrells, M. E., Baum, M., Wolters, P., and Powell, W. 2000. Assessment of genotypic varation among cultivated durum wheat based on EST-SSRs and genomic SSRs . Euphytica 119:39-43.

Feng, X., S. Saha, and K. M. Soliman, 1997. DNA fingerprinting in cotton using AFLP. Focus 19:11-12.

Goldstein, D. B., and D. D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J. Hered. 88:335-342.

Konieczny, A., and F. Ausubel. 1993. A procedure for quick mapping of Arabidopsis mutants using ecotype specific markers. Plant J. 4:403-410.

Liu, S., S. Saha, D. Stelly, B. Burr, and R. G. Cantrell. 2000. Chromosomal assignment of microsatellite loci in cotton. J. Hered. 91:326-332.

Moore, S. S., L. L. Sargeant, T. J. King, J. S. Mattick, M. Georges, and J. S. Hetzel. 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10:654-660.

Percival, A. E., J. F. Wendel, and J. M. Stewart. 1999. Taxonomy and germplasm resources. In: Smith, C. W., and J. T. Cothren (eds.) Cotton: origin, history, technology, and production. John Wiley & Sons, Inc., New York, pp. 33-64.

Qureshi, S. N., M. Altaf-Khan, J. N. Jenkins, C. L. Brubaker, U. K. Reddy, and S. Saha, 2001. A genetic linkage map of SSR markers based on *Gossypium nelsonii* and *G. australae* F2 population. Plant and Animal Genome Meeting IX, Jan. 13-17. San Diego, California. Abst. Pp. 219.

Rafalski, J. A, J. M. Vogel, M. Morgante, W. Powell, C. Andre, and S. V. Tingey. 1996. Generating and using DNA markers in plants. In: Birren, B. and E. Lai (eds.) Non-mammalian genomic analysis. A practical guide. Academic Press, New York, pp. 75-134.

Reddy, O. U. K., T. D. Brooks, K. M. El-Zik, and A. E. Pepper, 2000. Development and use of PCR-based technologies for cotton mapping. In P. Dugger and D. Richter (Eds.) Proc. Beltwide Cotton Research Conferences. Jan. 4-8, San Antonio, TX. National Cotton Council, Memphis, TN. 483 pp.

Reddy, O. U., A. E. Pepper, I. Abdurakmonov, S. Saha, J. Jenkins, T. Brook, and K. M. El-Zik. 2001. New dinucliotide and trinucliotide microsatellite marker resources for cotton genome research. J. Cotton. Sci. 5:103-113.

Reinisch, A. J., J. Dong, C. L. Brubaker, D. M. Stelly, J. F. Wendel, and A. H. Paterson. 1994. A detailed RFLP map of cotton, *Gossypium hirsutum X G. barbadense*: Chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829-847.

Shappley, Z. W., J. N. Jenkins, W. R. Meredith, and J. C. McCarty, Jr. 1998. An RFLP linkage map of upland cotton, *Gossypium hirsutum* L. Theor. Appl. Genet. 97:756-761.

Stewart, J. McD. 1995. Potential for crop improvement with exotic germplasm and genetic engineering. In: G. A. Constable and N. W. Forrester (Eds.). Challenging the Future: Proceedings. World Cotton Res. Conf.-1. CSIRO, Melbourne. pp. 313-327.

Tanksley, S. D., and S. R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066.

Ulloa, M., R. G. Cantrell, R. G. Percy, E. Zeiger, and Z. Lu. 2000. QTL analysis of stomatal conductance and relationship to lint yield in an interspecific cotton. J. Cotton Sci. 4:10-18.

Ulloa, M., W. R. Meredith Jr, Z. W. Shappley, and A. L. Kahler. 2002. RFLP genetic linkage maps from four $F_{2.3}$ populations and a joinmap of *Gossypium hirsutum* L. Theor. Appl. Genet. 104:200-208.

Vos, P., R. Hogersn, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Freijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. Nucl. Acids Res. 23:4407-4414.

Walbot, V. and L. S. Dure III. 1976. Developmental biochemistry of cotton seed embryogenesis and germination: 7. Characterization of the cotton genome. J. Mol. Biol. 101:503-536.

Weber, J. L. and P. E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44:388-396.

Yu, J., and R. J. Kohel. 1999. Update of the cotton genome mapping. In P. Dugger and D. Richter (Eds.) Proc. Beltwide Cotton Research Conferences. Jan. 3-7, Orlando, FL. National Cotton Council, Memphis, TN. 439 pp.

Yu, J., P. Yong-Ha, G. R. Lazo, and R. J. Kohel. 1998. Molecular mapping of the cotton genome: QTL analysis of fiber quality properties. In P. Dugger and D. Richter (Eds.) Proc. Beltwide Cotton Research Conferences. Jan. 5-9, San Diego, CA. National Cotton Council, Memphis, TN. 485 pp.

Zabeau, M., and P. Vos. 1993. Selective restriction fragment amplification: A general method for DNA fingerprinting. European Patent Application 92 402 629.7. Publ. EP 0 534 858.

Zhao, X, Y., Si, R. E. Hanson, C. F. Crane, H. J. Price, D. M. Stelly, J. F. Wendel, and A. H. Paterson 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Res. 8:479-492.